

## Authors' response to Anonymous Referee #1

---

We thank the reviewer for their time and comments that helped us to improve the paper. Responses to individual comments are provided below point-by-point. We also paste the text to reflect the changes to the manuscript.

---

**Comment 1:** The authors tend to apply the HIDRA model for business forecasting. However, the input for the model contains the atmospheric data in the future. Please clarify the datasource in the business forecast process.

**Response:** Thank you for pointing this out. As is usual in supervised learning, HIDRA was trained on past atmospheric model forecasts and sea level data. The HIDRA training data is composed of two parts: (a) the atmospheric part consists of past forecasts from a single member of ECMWF model ensemble, and (b) the sea level training data consists of tidal and residual sea levels from 24 hours prior to the forecast obtained from observations in Koper. For operational forecast, starting at time  $t_0$ , the following datasets will be employed:

- The atmospheric data: ECMWF ensemble forecast (the same ECMWF product as in training data) for time interval  $[t_0, t_0 + 72 \text{ hours}]$ . This data is available at the time of the forecast from ECMWF operational service.
- Sea Level data: tidal and residual data for the time interval  $[t_0 - 24 \text{ hours}, t_0]$ . This data is also available at the time of the forecast, assuming that the tide gauge in Koper works as expected, which is a reasonable assumption, given the redundancy in the design (the gauge consists of three independent bottom-mounted pressure gauges).

All the data required for HIDRA forecasting will be available for operational forecasts every day. To make this clear, the following text in the Conclusion,

*“HIDRA outperforms the current operational NEMO setup and is therefore an appropriate candidate for Agency’s operational pipeline.”*

was re-written into:

*“HIDRA outperforms the current operational NEMO setup and is therefore an appropriate candidate for Slovenian Environment Agency’s operational pipeline. HIDRA integration should be straightforward since ECMWF ensemble predictions and tide gauge sea level data are available at the Agency every day in real time for operational forecasting.”*

**Comment 2:** The authors declare “Extending the historical horizon beyond 24 hours did not significantly affect the prediction accuracy”. Please give a concise description of how to find the trade-off between the model forecast accuracy and the computing resource.

**Response:** To improve the insight into the trade-off, we conducted an additional experiment in which HIDRA was re-trained and tested for different values of historic horizon  $T_{\min}$ . The results are shown in Table 1. We find that increasing the historic horizon beyond 24h does not yield measurable improvements in prediction accuracy while increasing the number of parameters in the lower layers and negatively impacting the computational performance. For this reason we decided to set  $T_{\min} = 23$ .

	MAE [cm]	RMSE [cm]	Bias [cm]	Likelihood	CPU time [s]
<b>Overall</b>					
HIDRA <sub>12</sub>	5.3	7.0	<b>-0.2</b>	0.0440	<b>0.17</b>
HIDRA <sub>24</sub>	4.9	<b>6.4</b>	-0.4	<b>0.0470</b>	0.19
HIDRA <sub>36</sub>	<b>4.8</b>	<b>6.4</b>	-0.9	0.0455	0.21
HIDRA <sub>48</sub>	<b>4.8</b>	<b>6.4</b>	-0.6	0.0438	0.23
<b>Storm surge events</b>					
HIDRA <sub>12</sub>	11.7	13.9	-11.2	0.0220	<b>0.17</b>
HIDRA <sub>24</sub>	<b>10.3</b>	12.9	-9.3	<b>0.0253</b>	0.19
HIDRA <sub>36</sub>	10.7	13.2	-9.8	0.0245	0.21
HIDRA <sub>48</sub>	10.4	<b>12.8</b>	<b>-9.1</b>	0.0251	0.23

Table 1: HIDRA performance for different historic horizons in terms of mean absolute error (MAE), root mean squared error (RMSE) and model bias. CPU execution time (on a single core) per example is also reported. Performance for the atmospheric tide is provided as reference.

This experiment is now described along with the Table 1 in a new section in the manuscript (Section 4.1.1: Influence of the historic horizon):

*”We first analyze the influence of the HIDRA historic horizon defined by the parameter  $T_{\min}$  (see Section 3). Table 1 summarizes the performance of HIDRA with  $T_{\min} \in \{11, 23, 31, 47\}$ , which translates to historic horizons of 12, 24, 36 and 48 hours. Increasing the historic horizon from 12 to 24 hours significantly improves the prediction accuracy (9% reduction in RMSE error), however, further increases of the historic horizon (i.e., to 36 or 48 hours) do not show measurable benefits. Note that the execution time increases with the length of the historic horizon due to a substantial increase of parameters on the input layer. For this reason, we use a historic horizon of 24 hours ( $T_{\min} = 23$ ) as the best trade-off in the remaining analysis.”*

**Comment 3:** In Equation (1), there is a “20” on the “sum” signal, which represents the different spatial position on the feature maps, which is confusing that where this value comes from. Please clarify the changes of the feature maps during the fore-propagation process in the Figure. 3, especially in Figure 3 (a). Such as marking the size of the convolution kernel and the output size in the red boxes.

**Response:** Thank you for directing our attention to this issue. We agree that it is not immediately clear that the “20” in the spatial position sum follows from the input feature maps are of size  $4 \times 5$ . To address this, we now make the indexing in equation (1) explicit. In particular, the equation

$$\mathbf{f}_t = \text{ReLU} \left( \sum_{i=1}^{20} \mathbf{F}_t^{(i)} \mathbf{w}_t^{(i)} \right) \quad (1)$$

is now changed to

$$\mathbf{f}_t = \text{ReLU} \left( \sum_{i=1}^4 \sum_{j=1}^5 \mathbf{F}_t^{(i,j)} \mathbf{w}_t^{(i,j)} \right), \quad (2)$$

where  $(i, j)$  denotes the spatial coordinates of the feature map and spatial weights. We believe this is more precise and easier to follow.

Next, we address the comments regarding Figure 3a. We updated Figure 3 in the manuscript and made the annotations in the figure (see Figure 1) and the caption consistent with the text in the paper’s body. The Atmospheric Spatial Encoder now contains the standard markings of ResNet stages as well as the output feature map size for context. The previous caption of Figure 3

*“The proposed HIDRA architecture. A convolutional Atmospheric Spatial Encoder (ASE) extracts spatial atmospheric features from each time-step. Atmospheric and sea level temporal features are encoded by respective Temporal Encoder blocks, fused and passed to the fully-connected Residual Regression Block to predict the residuals along with their uncertainties. The trainable blocks are denoted by red color.”*

was changed to

*“The proposed HIDRA architecture. A convolutional Atmospheric Spatial Encoder (ASE) extracts spatial atmospheric features from each time-step. Atmospheric and sea level temporal features are encoded by respective Temporal Encoder blocks, fused and passed to the fully-connected Residual Regression Block to predict the residuals along with their uncertainties. With  $n$  we denote the number of filters or units of the block. The trainable blocks are colored red. The structure of the bottleneck blocks used in the ASE is presented in Figure 4.”*

Furthermore, we present the structure of the Bottleneck blocks in an additional figure (see Figure 2), which also details the convolutional parameters (kernel sizes, stride) and output feature map sizes.

**Comment 4:** Line 298: results  $\rightarrow$  result

**Response:** Thank you for for spotting the grammar mistake. The manuscript was updated as suggested.

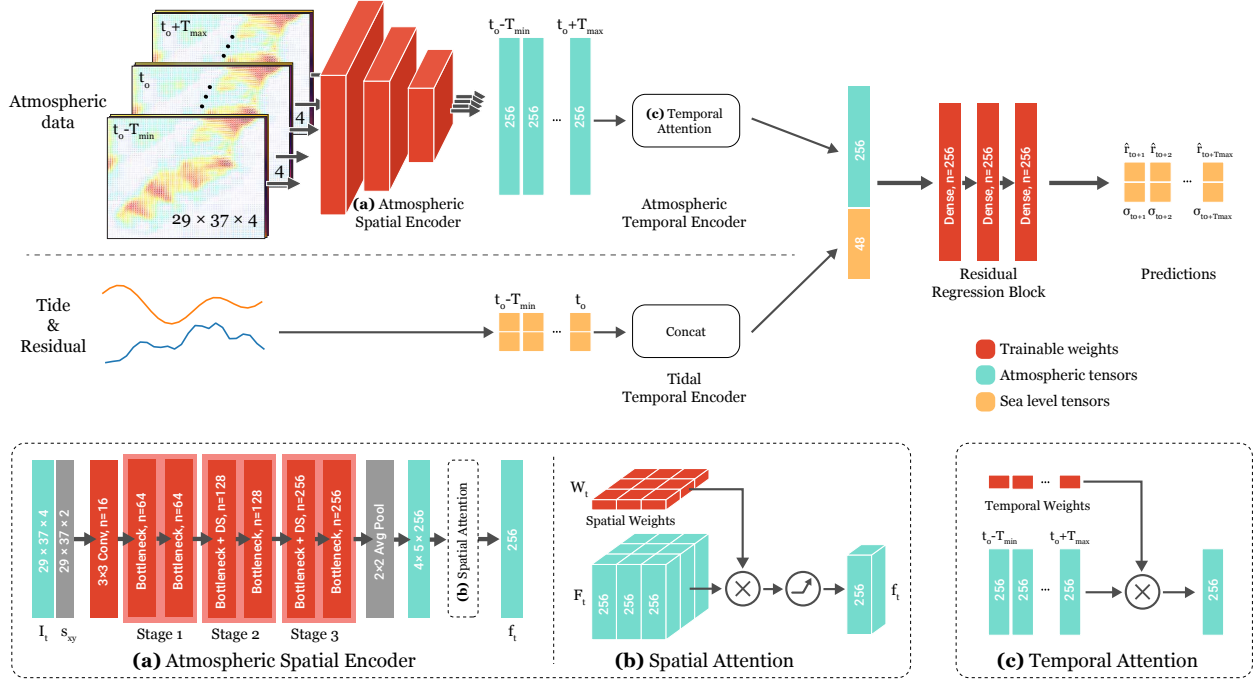


Figure 1: The proposed HIDRA architecture. A convolutional *Atmospheric Spatial Encoder* (ASE) extracts spatial atmospheric features from each time-step. Atmospheric and sea level temporal features are encoded by respective Temporal Encoder blocks, fused and passed to the fully-connected *Residual Regression Block* to predict the residuals along with their uncertainties. With  $n$  we denote the number of filters or units of the block. The trainable blocks are colored red. The structure of the bottleneck blocks used in the ASE is presented in Figure 2.

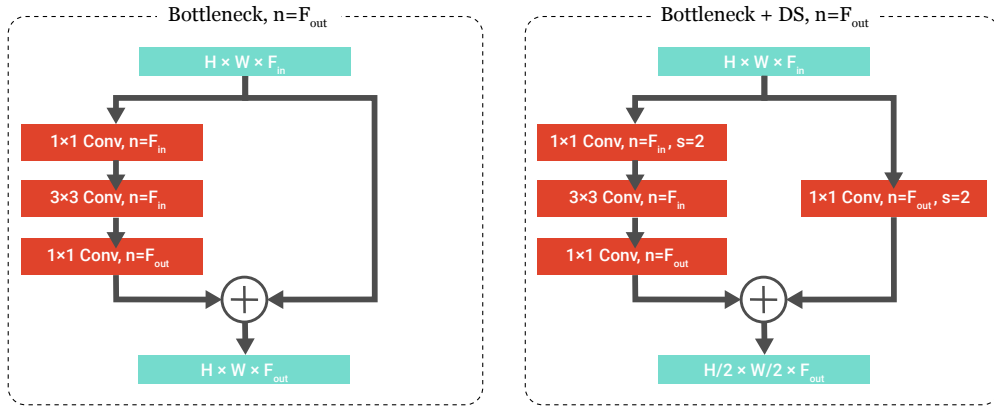


Figure 2: Structure of the bottleneck blocks used in the Atmospheric Spatial Encoder (Figure 1). The bottleneck block takes a feature map with depth  $F_{in}$  as the input and outputs a feature map with depth  $F_{out}$ . A regular bottleneck block (left) retains the spatial dimensions of the feature maps, while the downsampling (DS) bottleneck block (right) uses strided convolutions to reduce the spatial dimensions in half. We denote the number of convolutional filters by  $n$  and the stride parameter by  $s$ .